



Report on planned modifications

Team MaLi



Overview - Problems

- inappropriate translations, like "black mass" instead of "dark matter"
- Compound words often are translated incorrectly.
- There are a lot of translations with incorrect grammar tense
- Abbreviations are translated sometimes incorrectly.
- extra words appear
- Content words are missing



Solutions

Replace BPE with “wordpieces”

- several problems regarding BPE splitting

1. ID: 530

2. *Ungaris leiti, et peaaegu 96% lampidest on ohtlikud.*

3. *In Hungary, nearly 96% of the lights were found to be hazardous.*

4. *In Hungary, almost 96% of sheep were found to be dangerous.*

5. (1) Incorrect words, sense: 'sheep' instead of 'lights'

- → wordpiece
 - Variant of BPE
 - Final vocabulary size is fixed

Back translation

3. *They therefore postulate the existence of an invisible dark matter that only interacts with the rest of matter by its gravitation, and consists of very stable, and therefore very light, particles.*
4. *They are therefore exposed to the presence of an invisible black mass that is in interaction with other matter only through its gravity, and contains very stable, so very mild particles.*

- Improve quality/fluency
- Automatic translation of the monolingual target text → additional parallel training data
- No changes to neural network architecture → easy to apply

Hyperparameters – Beam Search

2. 2. *Klõpsake käsku Isikupärasta.*

3. 2. *Click personalize.*

4. *Click on your orders.*

- Could help to choose more appropriate words
- Change beam from default 5 to 10
- Involve candidates which are still close to the „best one“



Transformer Model

- Replaces recurrent layers with self-attention
- Connects all positions with a constant number of sequentially executed operations
- Dependencies between distant positions
- Requires less time to train